

Sigla Asignatura	BDY7101	Nombre de la Asignatura	BIG DATA
Nombre del Recurso Didáctico	1.1.12.Act_Sqoop		
Unidad de Aprendizaje N° 1	Ecosistema Hadoop		
Unidades de Competencia	Administra el volumen, la diversidad y complejidad del almacenamiento de datos para gestionar y extraer valor de acuerdo a los requerimientos de la organización.		

Actividad 1.1.12: Apache Sqoop

Descripción

En esta actividad podrán practicar con dos de las herramientas más populares del ecosistema Hadoop, Sqoop (SQL-to-Hadoop) y Hive (también Impala). Más específicamente, utilizando la ejecución por línea de comandos de Sqoop, deberán importar tablas a HDFS desde una base de datos relacional, que luego serán importadas en Hue al MetaStore de Hive (e Impala) para ser consultadas.

Instrucciones

1. Creación de la base de datos.

Descargue desde el sitio del curso la base de datos **movielens**, contenida en el archivo movielens.sql. A continuación acceda a través de la terminal a MySQL mediante el siguiente comando: **mysql -uroot -pcloudera**. Una vez dentro, cree una base de datos con nombre movielens: **create movielens** y luego importe el contenido del archivo descargado, mediante el comando **source filename.sql**. Para asegurarse que todo está correcto, ejecute cualquier consulta sobre la tabla movies.

2. Importación a HDFS.

Usando Sqoop, importe la tabla **movies** utilizando el siguiente comando:

```
sqoop import --connect jdbc:mysql://localhost/movielens --username root --password cloudera --fields-terminated-by '\t' --table movies
```

Verifique que el resultado sea correcto, primero revisando la carpeta **movies** en HDFS (**hadoop fs -ls movies**) y luego haciendo display del contenido del archivo resultante (**hadoop fs -cat movies/part-m-00000**).

Repita el proceso anterior para las tablas **ratings** y **users**.

Recursos

Usar las presentaciones del curso para apoyarse en las respuestas.

Criterio de evaluación

1. Carga la tabla **movies**. (3 puntos)
2. Carga la tabla **ratings**. (2 puntos)
3. Carga la tabla **users**. (2 puntos)